Integrative Biology

View Online
Dynamic Article Links

Cite this: DOI: 10.1039/c0ib00070a

www.rsc.org/ibiology

FRONTIER

# DNA assembly for synthetic biology: from parts to pathways and beyond†

**Tom Ellis,*[ab] Tom Adie[ac] and Geoff S. Baldwin*[ac]**

***Received 30th July 2010, Accepted 14th December 2010***
**DOI: 10.1039/c0ib00070a**



The assembly of DNA from small fragments into large constructs has seen significant recent development, becoming a pivotal technology in the ability to implement the vision of synthetic biology. As the cost of whole gene synthesis is decreasing, whole genome synthesis at the other end of the spectrum has expanded our horizons to the prospect of fully engineered synthetic cells. However, the recently proven ability to synthesise genome-scale DNA is at odds with our ability to rationally engineer biological devices, which lags significantly behind. Most work in synthetic biology takes place on an intermediate scale with the combinatorial construction of networks and metabolic pathways from registries of modular biopart components. Implementation for rapid prototyping of engineered biological circuits requires quick and reliable DNA assembly according to specific architectures. It is apparent that DNA assembly is now a limiting technology in advancing synthetic biology. Current techniques employ standardised restriction enzyme assembly protocols such as BioBricks™, BglBricks and Golden Gate methods. Alternatively, sequence-independent overlap techniques, such as In-Fusion™, SLIC and Gibson isothermal assembly are becoming popular for larger assemblies, and *in vivo* DNA assembly in yeast and bacillus appears adept for chromosome fabrication. It is important to consider how the use of different technologies may impact the outcome of a construction, since the assembly technique can direct the architecture and diversity of systems that can be made. This review provides a critical examination of recent DNA assembly strategies and considers how this important facilitating aspect of synthetic biology may proceed in the future.

## Introduction

Synthetic biology aims to apply the engineering principles of abstraction and characterisation to the design of biological systems.[1] In essence each DNA-encoded component or 'part', such as a promoter or open reading frame (ORF), can be considered in isolation. Robust characterisation of these individual parts means that they can be combined to produce new pathways and devices that give a predictable response, such as the expression of a protein in a host cell (chassis), to a specified level, under defined growth conditions. The abstraction envisioned by synthetic biology

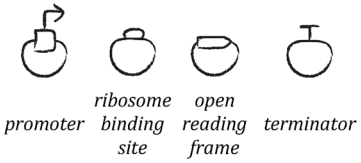
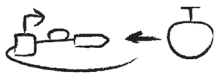
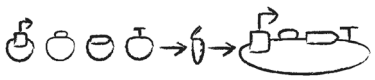
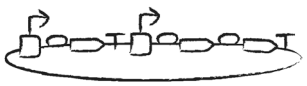
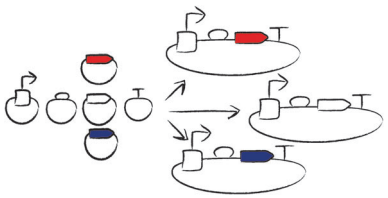
[a] *Centre for Synthetic Biology and Innovation, South Kensington, Imperial College London, SW7 2AZ, UK*
[b] *Department of Bioengineering, South Kensington, Imperial College London, SW7 2AZ, UK. E-mail: t.ellis@imperial.ac.uk; Fax: +44 (0)20-7594-7615*
[c] *Division of Molecular Biosciences, South Kensington, Imperial College London, SW7 2AZ, UK. E-mail: g.baldwin@imperial.ac.uk; Fax: +44 (0)20-7594-5288*

† Published as part of an iBiology themed issue on Synthetic Biology: Guest Editor Professor John McCarthy.

### Insight, innovation, integration

The development of synthetic biology requires integrating a rational engineering approach into the design of biological systems and is dependent on the robust characterisation of biological parts to assemble them into functional devices. The means by which DNA is assembled into devices is of pressing importance, since it is currently a limiting technology in the engineering of biological systems. This review provides a critique of current DNA assembly technologies and their applicability to synthetic biology as well as providing an insight into how the different methods affect the architecture and diversity at different scales of assembly.

**Table 1** An illustration of the terms used in DNA assembly for synthetic biology. Sequential and hierarchical assembly pathways produce constructs with a defined order of parts but this can be difficult to achieve using current one-step parallel methods, or when combinations of parts are used in order to yield a library of final constructs. ORF = Open Reading Frame

| | | | |
|---|---|---|---|
| ***Part*** DNA which performs a defined function in a genetic circuit and is compatible with an assembly technique |  | ***Ordered*** Assembled in a defined order |  |
| ***Gene*** An ORF along with all regulatory elements required for successful expression |  | ***Parallel*** One-step simultaneous assembly |  |
| ***Pathway*** A group of genes or operons (multi-ORF genes) which may perform related functions |  | ***Combinatorial*** Assembled with any number of 'alternative' parts at any number of defined positions; used in *e.g.* library cloning |  |

is to construct increasingly complex systems from DNA-encoded parts; combining parts to produce genes, linking genes to make pathways and devices, and finally arranging these together to create synthetic chromosomes and genomes (Table 1). One key aspect of moving up the levels of abstraction is better modelling and design capabilities (reviewed in this issue by MacDonald *et al.*[2]) and the other is the number and quality of characterised parts and devices. So although it is technically possible to synthesise at the whole genome scale,[3] our ability to rationally engineer biology is limited.[4] An ability to reliably assemble and test DNA components in a high throughput manner is so essential to this advancement that the limit of what synthetic biology can achieve is becoming determined by our ability to physically assemble DNA.

One of the foundational advances of synthetic biology was the BioBrick™, a DNA unit with standardised flanking sequences that enabled assembly to be achieved by a cheap, simple and standardised restriction/ligation method.[5,6] With BioBricks™ it became possible to store pots of modular biological parts that could be shared and easily assembled in different combinations by a vibrant community.[6,7] A decade on from the inception of standardised assembly, synthetic biology has produced a plethora of interesting and useful devices and pathways[8–12] and is poised to create more complex biological systems.[4,13,14] However, even with the simplicity of standardised assembly, the DNA assembly process is still a limiting factor for most laboratories, hampering rapid prototyping of many devices. Furthermore, the BioBrick™ approach becomes laborious when applied to

**Tom Adie**

*Tom Adie is a PhD candidate in the Centre for Synthetic Biology and Innovation at Imperial College, London, having obtained his Bachelors (Biology) and Masters in Research (Systems & Synthetic Biology) at the same institution. His research interests include DNA construction and cloning, directed evolution, automation and anything that will increase the efficiency of research in molecular biology.*

**Geoff S. Baldwin**

*Dr Geoff Baldwin received his PhD in Biochemistry working on DNA–protein interactions. He then went on to work in the laboratory of Prof. Steve Halford FRS (University of Bristol), before being awarded a David Phillips Research Fellowship from the BBSRC. His research has focused on understanding the molecular mechanisms of DNA damage recognition and repair by a number of specific DNA repair enzymes. More recently, Dr Baldwin's expertise on molecular enzymology has led to an interest in synthetic biology where his research focuses on the design of genetic circuits for the directed evolution of new bioparts.*

an intermediate assembly scale involving several genes and regulatory elements. New methods of assembly now need to be designed and implemented by the community, which can also be scaled to genome assembly.[15] Advance through this crucial stage in synthetic biology will be aided by a number of recently published and anticipated new techniques summarised in Fig. 1 and Table 2 and assessed in this article. These methods differ in both mechanism and scale, offering the user self-assembly of many parts in a single reaction (parallel assembly), giving constructs with a pre-defined physical arrangement (ordered assembly), or allowing multiple versions of parts to be used simultaneously (combinatorial assembly – see Table 1). The challenge for synthetic biology is to develop standardised assembly methods allowing work at all levels of abstraction – genes, pathways and genomes – and to clearly understand the context dependencies when parts are physically placed next to other parts. This review provides a critical overview of how the different assembly methods can be applied to different scales of assembly and how the inherent features of each method relate to the overall architecture and diversity of the end product.

## Parts to genes

A functional gene consists of a promoter, a translation start site (the RBS in prokaryotes), the protein coding ORF and lastly a terminator (Table 1). Other parts such as upstream and downstream regulatory elements, introns and RNA folding motifs also exist but are not essential. Assembly of a gene from its individual parts functionally requires **ordered** assembly; clearly placing a promoter after an ORF will not be worthwhile. Another important factor at this smallest scale is that very little DNA within a gene is non-functional, especially in prokaryotes, so 'scar' sequences – bases left behind by some assembly methods – are undesirable as they often affect how parts function together. Finally, the close proximity of parts at this scale brings up the issue of context-dependency. It is well established in biology that the behaviour of many DNA elements are context dependent, influenced by their immediate flanking sequences, and sometimes by distant ones. At the parts-to-genes level this is acutely illustrated by the RBS, the short RNA sequence that controls translation initiation.[16] Although the core sequence of an RBS is only 6 bases long it is always required to be exactly before the beginning of the ORF; and furthermore, flanking sequences ~50 bases around the RBS affect its efficiency, meaning the initial sequence of the ORF part will always modulate the RBS part.[16]

BioBrick™ assembly at this level has been particularly successful as the construction technique is sequential (and so inherently ordered; Table 1) and the number of parts needed to construct a functional gene is low. The major downside of the BioBrick™ approach is that the same 8 bp scar sequence is found at every junction. The presence of this scar sequence is unacceptable at certain positions, notably the RBS, meaning that alternative assembly methods must be used in cases where context-dependency is a problem. The scar is also problematic when assembling fusion proteins as it encodes an in-frame stop codon. To address these issues, revisions of the BioBrick™ standard have been introduced, starting with two standards specifically designed to assist assembly of fusion proteins.[17,18] More recently, a standard called BglBricks has been described[19] that uses different sequences for assembly and leaves a smaller 6 bp scar. This encodes a simple glycine-serine motif in frame, making the method more amenable to protein fusions. BglBrick assembly also has the advantage of using highly efficient and commonly-used restriction enzymes whose recognition sequences are not blocked by the most common DNA methylases, Dam and Dcm.

Despite revisions and new standards, neither BioBrick™ nor BglBrick methods can assemble a scarless gene from parts and crucially cannot assemble every sequence of DNA as the use of restriction enzymes means that the sequences they use as recognition sites are forbidden within a part. However, scarless assembly without any 'forbidden sites' is possible using other methods – notably overlap extension polymerase chain reaction (OE-PCR). This method, initially described over twenty years ago,[20] uses chimeric PCR primers 40+ bases long to create homologous ends between different DNA molecules, the homology is then used to prime extension in a second round of PCR between the initial products, and is the basis of most routine gene fusion techniques[21–23] The sequences of the homologous primers direct which parts follow each other, allowing ordered assembly which can be done sequentially, or even as a parallel reaction. The method is fast and relies on PCR, which is rapidly becoming more tractable for producing long constructs thanks to the increasing fidelity and decreasing cost of DNA polymerases. It is also dependent on ordering custom oligonucleotide primers for each concatenation, which are decreasing in cost but are still expensive when scaled to library-sized batches.

At the parts-to-genes level OE-PCR has always been a good alternative to restriction-ligation assembly. Typically it has been used to assemble 0.5 to 5 kb genes fragment *in vitro*, which would then be ligated into plasmids using restriction methods. Recently, however, a number of groups have demonstrated that OE-PCR has the power to assemble not just a gene, but the whole plasmid. Circular Polymerase Extension Cloning (CPEC)[24] utilises a single cycle of PCR without primers to circularise a gene into a linearised plasmid. The method is dependent on 20–25 bp overlap sequence at the ends of the

**Tom Ellis**

*Dr Tom Ellis is a lecturer in Synthetic Biology at Imperial College, London and researches at the Centre for Synthetic Biology and Innovation. Dr Ellis received his PhD in Pharmacology at Cambridge University before working in the laboratory of James J. Collins at Boston University, USA. His group does experimental research in synthetic biology, focusing on genetic engineering, genome design, the function of DNA sequence, standards in biological engineering, extreme life and the origin of cells.*

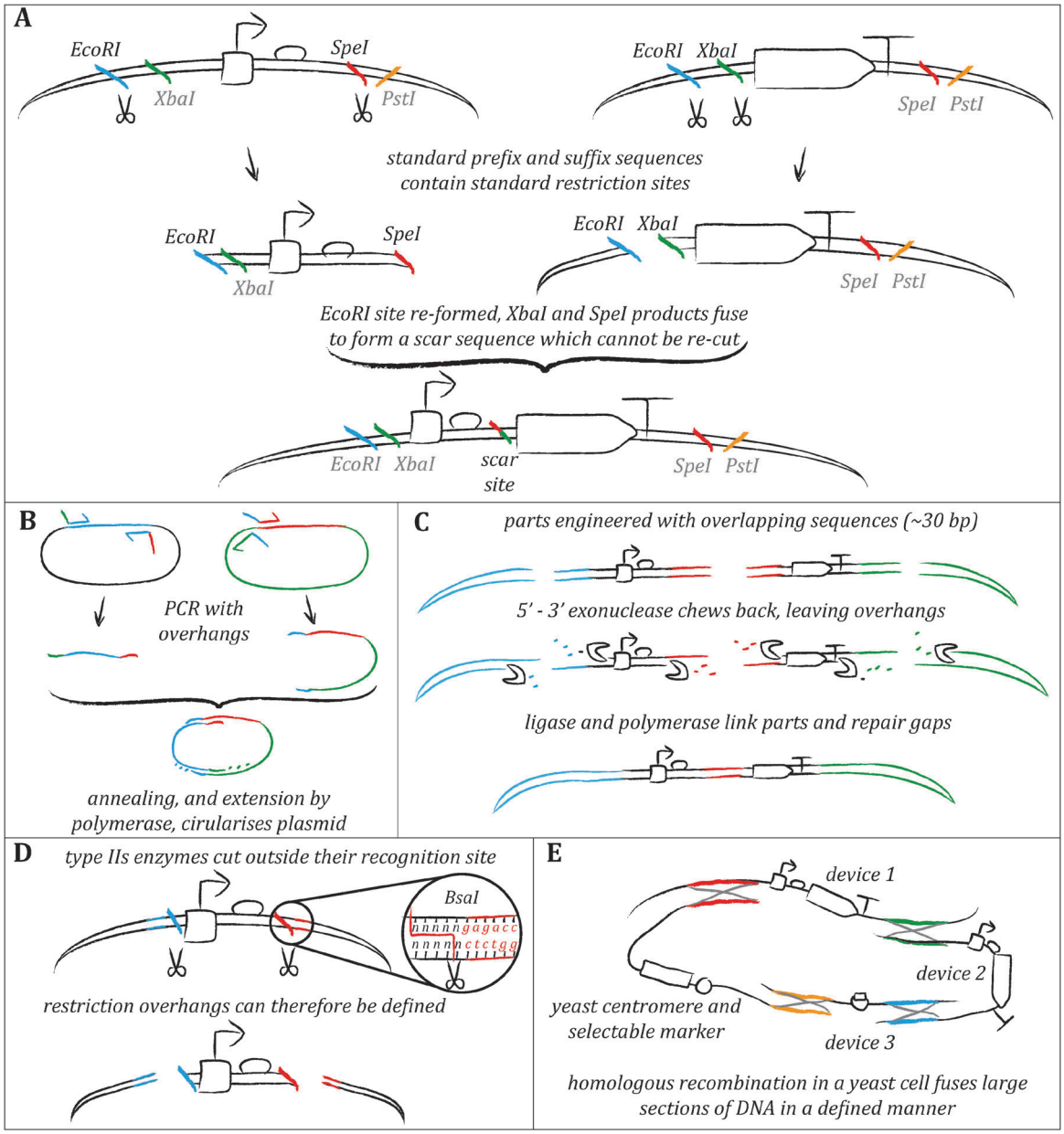



**Fig. 1** Overview of exemplar DNA assembly techniques. **A** shows BioBricks™ cloning of parts *via* restriction digestion and ligation of the resultant compatible sticky ends, re-forming the prefix sequence and leaving a 'scar' between parts. BglBricks swaps XbaI, SpeI and PstI with BglII, BamHI and XhoI respectively; thus the resultant scar does not include a stop codon or frameshift. **B** shows OE-PCR, where complementary overhangs attached to DNA targets anneal during PCR assembly and can be circularised into a vector by extension. **C** shows Gibson isothermal assembly of parts, where parts are synthesised to overlap by 30+ bp and their ends are processed and fused together using an exonuclease, a ligase and a polymerase. **D** illustrates Golden Gate assembly where type IIs enzymes cut outside their recognition site to excise parts with 4-base overhangs that are defined arbitrarily. With careful selection of compatible overhangs, such parts can perform parallel assembly into a defined order. **E** shows transformation-associated recombination (TAR) cloning in yeast, where parts including a yeast centromere and a selectable marker are inserted into yeast and naturally recombine in parallel through overlapping sequence to build a complete construct.

gene which are homologous to the ends of a linearised vector. The single PCR cycle with a high-fidelity proofreading DNA polymerase results in a small amount of nicked circular gene-plus-vector product that can be selected for with surprising efficiency by transforming into bacteria. With this high efficiency, CPEC may find particular usage in the combinatorial assembly often used in synthetic biology. Indeed, the authors demonstrate that both combinatorial library cloning and multi-part parallel assembly reactions are possible, with more cycles driving more complex reactions to completion. A recent similar technique[25] improves on CPEC by showing that the plasmid need not be linearised before assembly as long as the reaction products are digested with the methylation-sensitive *DpnI* restriction enzyme, which destroys any original circular DNA plasmid not produced by PCR, selecting only for assembled products. Together these two techniques offer a

**Table 2** Current DNA assembly techniques suitable for synthetic biology. For descriptions of the techniques see the main text. RE = Restriction Endonuclease; Square = Technique applicable in relevant context; Circle = Technique applicable but requiring specific considerations

| Method | *Mechanism* | Sequential | Parallel | Combinatorial | Parts - Genes | Genes - Pathways | Pathways - Genome |
|---|---|---|---|---|---|---|---|
| **BioBricks™** | Type II RE | ■ | | ■ | ■ | ■ | |
| **BglBricks** | Type II RE | ■ | | ■ | ■ | ■ | |
| **Pairwise Selection** | Type IIs RE | ■ | | ○ | | ■ | ■ |
| **Golden Gate** | Type IIs RE | ■ | ○ | ■ | ■ | ■ | |
| **InFusion™** | Overlap | ■ | ○ | ■ | ■ | ■ | |
| **Isothermal Assembly** | Overlap | ○ | ■ | ○ | | ■ | ■ |
| **SLIC** | Overlap | ○ | ■ | ○ | | ■ | |
| **USER** | Overlap | ○ | ■ | ○ | ■ | ■ | |
| **OE-PCR (incl. CPEC)** | PCR with Overlap | ■ | ■ | ○ | ■ | ■ | |
| ***Bacillus* Domino** | Recombination | ■ | | | | ○ | ■ |
| **Yeast TAR** | Recombination | | ■ | ○ | ○ | ■ | ■ |



powerful and fast new angle to OE-PCR, allowing parts-to-genes-to-plasmid assembly without any ligation reactions and using equipment and enzymes common to most labs. However, like all PCR methods, assembly is problematic when sequences contain many repeats or are GC-rich.

Perhaps the biggest advance in the first decade of synthetic biology is not a research breakthrough but the rapid decline in the cost of commercial synthesis of genes.[26] Genes are typically synthesised by polymerase cycling assembly from pools of overlapping custom oligos,[23,27] which can also now be prepared in miniature scale using microarray technology[28,29] and increasingly do not need to be purified before use.[30] Direct custom order synthesis of gene-sized DNA constructs is often cost-effective compared to acquiring the expertise and equipment required to assemble DNA from parts or traditional cloning techniques. Synthesis is typically one gene at a time; the order and the parts are defined in the design and there are no scars or forbidden sites. The huge disadvantage this has for synthetic biology is that individual parts, like promoters and ORFs, cannot be combinatorially swapped within constructs in order to create gene libraries without a large increase in synthesis costs. At this point the cost-balance switches and it is more common to have a codon-optimised ORF synthesised, which can then be assembled into a device using combinatorial techniques.

Despite this reservation, the advantages of commercial synthesis are still strong enough that it is predicted to dominate part-to-gene level assembly in the near future.[26] A huge body of research is building to understand codon-optimisation, which will dramatically improve predictability of expression for ORFs.[31–33] In addition, tools now exist to rationally predict other modulators of expression such as the RBS calculator,[16] which uses RNA-folding models to output sequences that will trigger translation initiation at desired strengths. Tools like the RBS calculator and similar promoter calculators[34] will drive the move to custom synthesis of genes with defined expression levels. However, there will still always be a requirement to combinatorially exchange parts within these genes for others from part libraries, particularly in the case of the regulatory promoter, which not only drives expression levels, but typically also controls the logic of gene regulation.[35,36]

## Genes to pathways

While building genes from parts is the bedrock of synthetic biology, most research projects in the field operate at the higher scale of linking genes to construct pathways and devices. From iGEM projects building ingenious regulatory networks[12,37] to industrial synthesis of biofuels through metabolic

engineering,[38,39] assembly at this level is the key area that is currently challenging synthetic biology. Unlike the parts-to-genes level, the presence of scar sequences between genes and their physical order in a DNA construct may not dramatically affect the function of a pathway, unless it is arranged in an operon. Instead, the importance at this level is to be able to assemble genes **combinatorially** so that libraries of genes can yield massive diversity in possible pathways and devices.

As at the parts-to-genes level, BioBrick™ methods are suitable for creating small pathways and devices like operons, and are exceptionally useful for combinatorial assembly. However, the slow laboratory methodology and requirement of forbidden restriction sites does not favour scaling-up to larger assemblies. Attempts to automate this kind of assembly can help speed up the process, but will always be limited by the sequential nature of the technology. Furthermore, the nature of the scar site means that individual parts cannot be replaced once assembled. OE-PCR methods can also be applied to create operons and small pathways and devices, especially recent CPEC and related methods that can combinatorially assemble many plasmids in one go. However, these methods are also limited in their ability to scale up, both by plasmids becoming less efficient at larger sizes and by the error rate of PCR, which remains too erroneous for reliable perfect amplification of 10+ kb constructs.

The challenge of assembly at larger scales has led to the emergence of several new methods utilising Type IIs restriction enzymes. Although not routinely used in standard cloning, these enzymes have in the past found application in niche protocols such as those for assembling repetitive DNA[40–42] and were utilised by Kodumal *et al*. for synthetic assembly of a 32-kb gene cluster.[43] They cut outside their recognition sequence, allowing 4 bp overhangs to be customised and thus assembly can be scarless between neighbouring fragments designed to overlap by the appropriate 4 bp.[44,45] Recently this enzymatic property has been used in a new Golden Gate assembly method; a parallel one-pot, one-step 5 min technique to assemble seamless constructs.[46] DNA is inserted into an entry clone shuttle vector which provides the Type IIs recognition sequences immediately at both ends of the DNA pieces. Digestion then produces all the fragments for assembly which ligate in parallel where overhangs are complementary. While this method is proficient for assembly at the parts-to-genes level and for smaller pathways and devices, it can also be used to shuffle multiple fragments. This shuffling variation[47] uses PCR to add the 4 bp overlap sequences to the ends of each fragment, which are chosen to define the order of the fragments in the assembly reaction. Theoretically 4 bp overhangs allow 240 distinct overlaps, but the authors report base-pair mismatching at similar overhangs can result in incorrect assembly. Already a version of Golden Gate shuffling that does not require shuttle vectors, but instead adds Type IIs recognition sequences by PCR, has been proposed for synthetic biology.[48] The promise of ordered, parallel assembly with no scars, or just 4 bp scars, is definitely attractive for both parts-to-genes and genes-to-pathways assembly.

Interestingly Blake *et al*.[49] recently described a more complicated version of Type IIs cloning specifically designed to facilitate the assembly of large constructs from directly synthesised gene-length fragments. Their novel assembly method requires attachment of ~65 bp standardised tags at both ends of the composite fragments, and the use of two cassette vectors that each contain two different non-functional antibiotic resistance markers. Type IIs recognition sequences in the tags allow the fragment to be cut and ligated into the vectors at which point the tags act as promoters for the antibiotic resistance markers, enabling stringent selection of vector containing the insert. Two plasmids containing insert fragments with 4 bp overlap can then be assembled together by digesting with a second set of type IIs enzymes so that inserts are excised with tags only at one end of each. These two pieces are then ligated to become a single larger insert that goes into a second vector with two different antibiotic resistance markers that are selected for. Repeating the pairwise selection cycle with 1 to 2 kb starting fragments can build a 91 kb assembly in only 6 rounds of transition between cassette vectors.

The use of tags containing Type IIs sites that double as promoters for stringent antibiotic resistance markers allows this method to be done entirely in liquid culture, facilitating faster throughput using automation with a liquid handling robot. The forbidden site problem is circumvented using a clever twist whereby sequence-specific blocking oligos prevent CpG methylation of tag digest sites, *via RecA*-mediated binding to homologous sequences. This leaves them open for further digestion, unlike those in the assembly fragments which become protected. This oligo-based technology (RARE)[50] could well become adopted elsewhere in DNA assembly protocols that have a forbidden site requirement.

At the parts-to-pathways level, the alternative to restriction digestion methods are a class of techniques called overlap assembly methods. Like OE-PCR, overlap methods require DNA fragments for assembly to share 20+ bp of common sequence at ends that will be joined. This sequence is processed *in vitro* by enzymes that perform the assembly. In the case of well-known commercial kits for plasmid construction such as Gateway™ (Life Technologies), the enzyme is a specific recombinase and the overlap sequence guiding assembly must be specific for this.[51] This is undesirable for large assembly projects as the same sequence will be present at each junction and large parallel assemblies are impossible. A more promising commercial kit also for cloning into plasmids is In-Fusion™ (Clontech) which uses a proprietary enzyme mix to assemble any fragments with 15 bp sequence overlap.[52] It is particularly suited for parallel reactions with several DNA fragments and conceptually belongs to a class of overlap assembly techniques often called 'chew back and anneal'. These methods work by digesting back one strand of DNA from each exposed end of a fragment, leaving a single-stranded overhang that anneals with the complementary overhang of a fragment sharing the same overlap sequence. Those fragments may then be transformed directly or ligated together to give scarless assembly.

An established overlap method is USER (uracil-specific excision reagent) cloning which first requires PCR amplification of fragments using primers that incorporate at least one uracil.[53,54] Once incorporated into the sequence, the uracils are excised by uracil DNA glycosylase and the resultant abasic sites cleaved by an AP-lyase, leaving 3′ overhangs complementary to the original primer sequence. Such overhangs (which can be 7–8 bases, or as long as desired) can thus be produced and assembled

in a single ligase-independent method.[53] It should be noted that ligation is technically feasible with this method, since AP lyase enzymes leave a 5′ $PO_4$, while the excised oligonucleotide cannot re-ligate back in as it will possess a 3′ deoxyribose phosphate. Like other overlap methods, no assembly scars are left behind and parallel assembly is possible. The drawback to this method is that at least one thymidine is required near the end of the sequence (to be replaced with a uracil) so it is not truly 'sequence-independent'.

Several other ligation-independent cloning methods also exist,[55,56] and a sequence-independent variation has recently been described.[57] This method, SLIC (sequence- and ligation-independent cloning) utilises a slow chew back mechanism resulting from the 3′-5′ proofreading exonuclease activity T4 DNA polymerase exhibits in the absence of dNTPs. In this case, 30 min incubations create long overhangs of 30+ bp overlaps which anneal with homologous excised fragments. *RecA* is used to stabilise the annealed fragments and avoid *in vitro* ligation, instead relying on *in vivo* DNA repair of any gaps post-transformation. In a demonstration of its parallel capability, a single SLIC reaction was used to generate a ten-way assembly of 300–400 bp PCR fragments;[57] an impressive feat that could be applied to pathway assembly.

Though applicable to the pathway scale, most overlap assembly methods were originally described for plasmid construction experiments, rarely going beyond 15+ kb. This is not the case for three overlap methods recently published by the J. Craig Venter Institute.[58] These methods were intended to assemble whole genomes *in vitro* from directly synthesised 5 kb fragments designed to have 40 to 400 bp of overlapping sequence. The protocols are all chew back and anneal methods using cocktails of different enzymes, and the most impressive version requires a single 30 min incubation at one temperature. This 'Gibson' isothermal assembly method, using a high fidelity DNA polymerase, T5 exonuclease and *Taq* DNA ligase, is particularly attractive due to its simplicity and use of common lab enzymes.

In demonstrating these methods, Gibson *et al.* successfully assembled a complete synthetic 583 kb *M. genitalium* genome *in vitro* from four 100 kb+ fragments. They also showed their protocols to be efficient with fragments at the 2 kb scale and have recently adapted the enzyme ratios in their preparation to allow the method to be used to assemble genes directly from single-stranded 60-mer oligos that overlap by 20 bases, bypassing gene synthesis.[59] This latter variation of the protocol has allowed Gibson *et al.* to successfully assemble the entire 16.3-kilobase mouse mitochondrial genome from 600 overlapping 60-mers using only the Gibson isothermal assembly method at all stages, showing the power of this approach.[59] The technique is massively parallel and can assemble circular plasmids and bacterial artificial chromosomes (BACs) in single reactions. Their work clearly demonstrates a future assembly approach for large pathways and devices, where directly synthesised genes are parallel-assembled into huge constructs without having to consider the forbidden site requirement of restriction enzymes, or the error-rate of PCR over long DNA lengths. As with all overlap assembly methods the order of fragments is pre-determined by the sequence overlap between them. While this method has been ideal for re-synthesising a known genome, it remains to be used in the context of combinatorial assembly of pathways.

## Pathways to genomes

The future for synthetic biology will be integrating assemblies of pathways, devices and regulatory networks to build designer genomes for custom microbes. Research in this area is already active and beginning to bear fruit.[60] At the pathways to genome level, it will be crucial to have **parallel** assembly reactions, as sequential construction from fragments will be laborious for large assemblies. It is also becoming evident that the positioning of genes, pathways and devices on genomes, especially in bacteria, plays an important role in their regulation and the intracellular location of the products they encode,[61] and so the order of assembly will need to be defined.

Despite successfully demonstrating *in vitro* assembly of a genome using a chew back and anneal overlap technique, the J. Craig Venter Institute assembled the first complete synthetic genome *in vivo* by using *S. cerevisiae* yeast homologous recombination.[3,62,63] They developed a modified Transformation-Assisted Recombination (TAR) cloning protocol able to assemble the entire circular synthetic *M. genitalium* genome in one go from 25 pieces about 24 kb in size. TAR has been a common protocol for manipulating DNA in yeast for over a decade,[64] and like chew back and anneal methods it relies on overlapping sequence between fragments. These undergo homologous recombination during yeast spheroplast transformation, and this technique is predominantly used to incorporate gene- and pathway-sized DNA assemblies into specific sites in the yeast genome. By including a yeast artificial chromosome (YAC) replication sequence and a selective marker in one assembly fragment, assembly of a circular self-propagating construct is achieved.

Due to the fidelity of the native homologous recombination enzymes, assembly in yeast is inherently accurate. Yeast has the advantage of being tolerant to very large constructs; stable YACs as large as 2.3 Mbp have been reported.[65] The massively-parallel assembly is also rapid and Gibson *et al.* speculate that yeast may be able to assemble an entire genome from the same 110 synthesised ~5 kb fragments used for their *in vitro* isothermal method. Remarkably the assembly of 25 pieces into an entire genome was achieved with overlaps between fragments as short as 80 bp, and TAR cloning has been reported with overlap sequences as short as 40 bp.[66] Further research by the same authors has demonstrated the power of yeast as a vehicle for gene synthesis, with cells capable of assembling 1+ kb genes *in vivo* from overlapping 60-mer oligonucleotides.[67] Additionally, yeast is also tractable for pathway assembly, as demonstrated by the similar 'DNA Assembler' method developed concurrently by Shao *et al.*[68]

For genome-sized assembly, yeast is not the only cellular chassis; recently the complete assembly of chromosomes was also achieved in *Bacillus* by Itaya *et al.*[69] Building on a previous 'inchworm' method[70] that had been used to clone 3500 kb of *Synechocystis* DNA into *B. subtilis* from long 100 kb+ starting pieces, the authors developed a new 'domino method' utilising homologous recombination. Their method uses specific *B. subtilis* strains called Bacillus genome (BGM) vectors that

have integrated pBR322 sequences in their genomes. Assembly is sequential and requires the target chromosome to be broken down into 5 kb regions with significant overlap between each piece. Pieces to be assembled are cloned in *E. coli* alternately into two vector plasmids with different antibiotic resistance markers. Assembly begins by adding the first plasmid to *Bacillus* which recombines the assembly piece plus resistance marker into the BGM vector site of the genome. The process is then repeated for the second plasmid, and this time homologous recombination replaces the first resistance marker with the next part of the construct, and adds a second. Repeated transformation and alternate selection with the two antibiotics sequentially builds the desired chromosome in the *Bacillus* genome and this can be recovered by transforming a retrieval plasmid or BAC that excises the assembly by homologous recombination.

The domino method was used to assemble the complete organelle genome of the mouse mitochondria (16.3 kb), and modified to include more antibiotic markers in order to assemble the rice chloroplast genome (134.5 kb). This latter achievement is particularly of note, as this genome is highly-repetitive and would be difficult to assemble by other means. Although ordered, unfortunately both the domino and the previous inchworm methods are not rapid parallel methods but laborious, sequential 'wet-lab-heavy' processes. They also rely on long overlaps for recombination and cannot be used to assemble any sequences that already exist in the *Bacillus* genome. Interestingly, further development of their protocol now allows large DNA fragments on BACs to be transferred directly in culture from *E. coli* to *B. subtilis* in a culture mix method, without any purification.[71] Given the difficulties associated with handling large DNA fragments[72] this method could be beneficial to adopt elsewhere.



## Future developments

An idealised assembly method for synthetic biology must be suitable for combinatorial construction from standarised part libraries, have no forbidden site requirements, and allow for pre-determined order in the final product. Yet it would also need to be assembled rapidly in a parallel reaction, applicable for work at any scale and only leave scars between parts that can tolerate them. Currently, no single technique is capable of all of these and different methods have varying suitability to different levels of assembly, so often the most appropriate strategy involves using several techniques in tandem. At the parts-to-genes level, the future will be dominated by ever-cheaper direct synthesis and tools to predict expression, but synthetic biologists will still need assembly to exchange promoter sequences from part libraries and won't be able to tolerate scar sites unless the context-dependencies of these are fully understood or new 'insulator' parts can be designed to prevent part function being influenced by flanking sequence. A further challenge at this level will undoubtedly be placed by expanding synthetic biology beyond the simplest case of minimal bacterial genes and expanding the tool-set of regulatory parts to include enhancers, regulatory RNA and motifs essential in eukaryotic systems such as those involved into chromatin packing and remodelling.

At both the genes-to-pathways and pathways-to-genomes level, the recent advances in assembly described here offer promise for improved techniques to greatly aid progress in synthetic biology. In particular, the parallel overlap-based techniques have the advantage of no forbidden site requirements; an important feature as even with direct synthesis and codon-optimization there will always be instances where a sequence must be included in the final construct. Many overlap methods are also cross-compatible, as illustrated by assembly of the same fragments by either *in vitro* or *in vivo* methods by the J. Craig Venter Institute. The challenge with overlap methods is to introduce combinatorial assembly from pots of modules without loss of parallel assembly. Current parallel techniques work for synthesised fragments that overlap with neighbours or rely on using PCR with custom primers to add each overlap. These methods aren't scalable for future projects to create pathways built from gene libraries or when rearranging the order of genes; assembling 10 overlapping DNA fragments in every possible arrangement would require 90 versions of each primer.

We propose that an attractive way around these problems would be to develop a set of standardised overlap sequences compatible with several assembly techniques. Overlaps, perhaps 40 to 60 bp long, would be non-repeating scars between genes and could be designed to define the location of each gene, with *n* genes spaced along a chromosome with *n* different overlap sequences between. They would themselves be a set of standard parts and could feasibly double as insulators. With a standard set of PCR primers, these overlaps could easily be added to gene-length fragments having common prefix and suffix sequences, such as those in current synthetic biology libraries.

Undoubtedly, future DNA assembly protocols will be greatly aided by computational tools to aid design and to automate the liquid-handling of core methods (reviewed in detail in this issue by MacDonald *et al.*[2]). Research in assembly automation is active, in both the theoretical analysis of strategies[73,74] and the development of Computer Aided Design (CAD) software suites such as Clotho.[75] Already, software tools such as the RBS Calculator[16] are being used to direct design in many labs, and tools aiding new methods of assembly will be required where necessary. Life Technologies, for example, now offer online design tools for Gibson isothermal assembly and yeast-based DNA assembly-methods that they sell though GENEART as commercial kits. The next stage for design software should be inclusion of model-based understanding of the context-dependencies of neighbouring parts which will in turn inform the assembly strategy. Software tools will need to be designed to allow mix-and-match assembly workflows that use the multiple techniques. The Joint BioEnergy Institute (JBEI) have developed an in-house software package, *J5*, specifically for this purpose[76] and are able to link the software to direct automation of assembly by liquid handling robots.

Greater use of automated liquid handling in DNA assembly will increase throughput but another future direction to consider is the use of *in vivo* methods where engineered bacterial cells could be programmed to use methylases and restriction enzymes within the cell to select what DNA sequences to cut and rearrange. Already techniques exist to directly edit genomic DNA within cells; the recently developed multiplex-automated genomic engineering (MAGE) protocol

is an example of this.[77] With the cost of whole genome sequencing falling rapidly, verification of multiple DNA changes within cells is now becoming affordable. This is an important consideration as most assembly methods still rely on traditional single-read sequencing for verification at the end-point and at intermediate stages, and this form of sequencing is not decreasing in cost and time. To enable future DNA assembly techniques to be rapid and affordable, new methods to quickly sequence-verify multiple intermediates such as plasmid DNA constructs will be required.

## Acknowledgements

We'd like to thank Travis Bayer, James Chappell, Chris Hirst, Rui Rodrigues, Guy-Bart Stan, James Field, Rhys Algar, Richard Kitney and Paul Freemont at Imperial College, London's Centre for Synthetic Biology and Innovation for the useful discussions regarding standardised assembly. We would like to thank the EPSRC for funding.



## References

1 E. Andrianantoandro, S. Basu, D. K. Karig and R. Weiss, Synthetic biology: new engineering rules for an emerging discipline, *Mol. Syst. Biol.*, 2006, **2**, 2006.0028.
2 J. MacDonald, C. Barnes, R. I. Kitney, P. Freemont and G.-B. Stan, Computational design approaches and tools for synthetic biology, *Integr. Biol.*, 2010, DOI: 10.1039/c0ib00077a, this issue.
3 D. Gibson, J. Glass and C. Lartigue, *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome, *Science*, 2010, **329**, 52.
4 P. E. M. Purnick and R. Weiss, The second wave of synthetic biology: from modules to systems, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 410–22.
5 T. Knight, Idempotent Vector Design for Standard Assembly of Biobricks, *DSpace*, 2003, http://dspace.mit.edu/handle/1721.1/21168.
6 C. D. Smolke, Building outside of the box: iGEM and the BioBricks Foundation, *Nat. Biotechnol.*, 2009, **27**, 1099–102.
7 C. Goodman, Engineering ingenuity at iGEM, *Nat. Chem. Biol.*, 2008, **4**, 13.
8 S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold and R. Weiss, A synthetic multicellular system for programmed pattern formation, *Nature*, 2005, **434**, 1130–4.
9 A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church and J. J. Collins, Synthetic gene networks that count, *Science*, 2009, **324**, 1199–202.
10 D. K. Ro, E. M. Paradise and M. Ouellet, *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast, *Nature*, 2006, **440**, 940–3.
11 T. Danino, O. Mondragón-Palomino, L. Tsimring and J. Hasty, A synchronized quorum of genetic clocks, *Nature*, 2010, **463**, 326–30.
12 J. J. Tabor, H. M. Salis and Z. B. Simpson, *et al.* A synthetic genetic edge detection program, *Cell*, 2009, **137**, 1272–81.
13 A. S. Khalil and J. J. Collins, Synthetic biology: applications come of age, *Nat. Rev. Genet.*, 2010, **11**, 367–79.
14 T. K. Lu, A. S. Khalil and J. J. Collins, Next-generation synthetic gene networks, *Nat. Biotechnol.*, 2009, **27**, 1139–50.
15 P. A. Carr, DNA construction: homemade or ordered out?, *Nat. Methods*, 2010, **7**, 887–9.
16 H. Salis, E. Mirsky and C. Voigt, Automated design of synthetic ribosome binding sites to control protein expression, *Nat. Biotechnol.*, 2009, **27**, 946.
17 R. Grünberg, K. Arndt and K. Müller, BioBricks Foundation RFC 25: Fusion Protein (Freiburg) BioBrick assembly standard, *DSpace*, 2009, http://dspace.mit.edu/handle/1721.1/45140.
18 I. Phillips and P. Silver, BioBricks Foundation RFC 23: A New BioBrick Assembly Strategy Designed for Facile Protein Engineering, *DSpace*, 2006, http://dspace.mit.edu/handle/1721.1/32535.
19 C. Anderson, J. Dueber and M. Leguia, *et al.* BglBricks: A flexible standard for biological part assembly, *J. Biol. Eng.*, 2010, **4**, 1.
20 R. Horton, H. Hunt, S. Ho, J. Pullen and L. Pease, Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension, *Gene*, 1989, **77**, 61.
21 K. Cha-aim, T. Fukunaga, H. Hoshida and R. Akada, Reliable fusion PCR mediated by GC-rich overlap sequences, *Gene*, 2009, **434**, 43–9.
22 J. R. Chen, J. J. Lü and H. F. Wang, Rapid and efficient gene splicing using megaprimer-based protocol, *Mol. Biotechnol.*, 2008, **40**, 224–30.
23 M. Czar, J. Anderson, J. Bader and J. Peccoud, Gene synthesis demystified, *Trends Biotechnol.*, 2009, **27**, 63.
24 J. Quan and J. Tian, Circular polymerase extension cloning of complex gene libraries and pathways, *PLoS One*, 2009, **4**, e6441.
25 A. Bryksin and I. Matsumura, Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids, *BioTechniques*, 2010, **48**, 463.
26 R. Carlson, The changing economics of DNA synthesis, *Nat. Biotechnol.*, 2009, **27**, 1091–4.
27 A. S. Xiong, R. H. Peng and J. Zhuang, *et al.* Chemical gene synthesis: strategies, softwares, error corrections, and applications, *FEMS Microbiol. Rev.*, 2008, **32**, 522–40.
28 J. Tian, H. Gong and N. Sheng, *et al.* Accurate multiplex gene synthesis from programmable DNA microchips, *Nature*, 2004, **432**, 1050–4.
29 C. C. Lee, T. M. Snyder and S. R. Quake, A microfluidic oligonucleotide synthesizer, *Nucleic Acids Res.*, 2010, **38**, 2514–21.
30 A. Y. Borovkov, A. V. Loskutov and M. D. Robida, *et al.* High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides, *Nucleic Acids Res.*, 2010, **38**, e180.
31 N. Mitarai, K. Sneppen and S. Pedersen, Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization, *J. Mol. Biol.*, 2008, **382**, 236–45.
32 M. Welch, S. Govindarajan and J. E. Ness, *et al.* Design parameters to control synthetic gene expression in Escherichia coli, *PLoS One*, 2009, **4**, e7002.
33 M. Hamady, S. A. Wilson, J. Zaneveld, N. Sueoka and R. Knight, CodonExplorer: an online tool for analyzing codon usage and sequence composition, scaling from genes to genomes, *Bioinformatics*, 2009, **25**, 1331–2.
34 V. A. Rhodius and V. K. Mutalik, Predicting strength and function for promoters of the Escherichia coli alternative sigma factor, sigmaE, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2854–9.
35 T. Ellis, X. Wang and J. J. Collins, Diversity-based, model-guided construction of synthetic gene networks with predicted functions, *Nat. Biotechnol.*, 2009, **27**, 465–71.
36 R. S. Cox, M. G. Surette and M. B. Elowitz, Programming gene expression with combinatorial promoters, *Mol. Syst. Biol.*, 2007, **3**, 145.
37 C. Lou, X. Liu and M. Ni, *et al.* Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch, *Mol. Syst. Biol.*, 2010, **6**, 350, DOI: 10.1038/msb.2010.2.
38 J. M. Carothers, J. A. Goler and J. D. Keasling, Chemical synthesis using synthetic biology, *Curr. Opin. Biotechnol.*, 2009, **20**, 498–503.
39 S. K. Lee, H. Chou, T. S. Ham, T. S. Lee and J. D. Keasling, Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels, *Curr. Opin. Biotechnol.*, 2008, **19**, 556–63.
40 S. L. Berger, R. E. Manrow and H. Y. Lee, Phoenix mutagenesis: one-step reassembly of multiply cleaved plasmids with mixtures of mutant and wild-type fragments, *Anal. Biochem.*, 1993, **214**, 571–9.
41 K. A. Padgett and J. A. Sorge, Creating seamless junctions independent of restriction sites in PCR cloning, *Gene*, 1996, **168**, 31–5.
42 J. McDaniel, A. MacKay, F. Quiroz and A. Chilkoti, Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes, *Biomacromolecules*, 2010, **11**, 944.
43 S. J. Kodumal, K. G. Patel, R. Reid, H. G. Menzella, M. Welch and D. V. Santi, Total synthesis of long DNA sequences: synthesis

of a contiguous 32-kb polyketide synthase gene cluster, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15573–8.
44 W. Szybalski, S. C. Kim, N. Hasan and A. J. Podhajska, Class-IISrestriction enzymes–a review, *Gene*, 1991, **100**, 13–26.
45 S. M. Lippow, P. M. Aha, M. H. Parker, W. J. Blake, B. M. Baynes and D. Lipovsek, Creation of a type IIS restriction endonuclease with a long recognition sequence, *Nucleic Acids Res.*, 2009, **37**, 3061–73.
46 C. Engler, R. Kandzia and S. Marillonnet, A One Pot, One Step, Precision Cloning Method with High Throughput Capability, *PLoS One*, 2008, **3**, e3647.
47 C. Engler, R. Gruetzner, R. Kandzia and S. Marillonnet, Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes, *PLoS One*, 2009, **4**, e5553.
48 S. G. Peisajovich, A. Horwitz, O. Hoeller, B. Rhau and W. Lim, BioBricks Foundation RFC 28: A method for combinatorial multi-part assembly based on the Type IIs restriction enzyme AarI, *DSpace*, 2009, http://dspace.mit.edu/handle/1721.1/46721.
49 W. Blake, B. Chapman, A. Zindal, M. Lee, S. Lippow and B. Baynes, Pairwise selection assembly for sequence-independent construction of long-length DNA, *Nucleic Acids Res.*, 2010, **38**, 2594.
50 L. J. Ferrin and R. D. Camerini-Otero, Selective cleavage of human DNA: RecA-assisted restriction endonuclease (RARE) cleavage, *Science*, 1991, **254**, 1494–7.
51 Y. Sasaki, Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system, *J. Biotechnol.*, 2004, **107**, 233.
52 S. Sleight, B. Bartley, J. Lieviant and H. Sauro, In-Fusion BioBrick assembly and re-engineering, *Nucleic Acids Res.*, 2010, **38**, 2624.
53 H. H. Nour-Eldin, F. Geu-Flores and B. A. Halkier, USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories, *Methods Mol. Biol.*, 2010, **643**, 185–200.
54 C. Smith, P. J. Day and M. R. Walker, Generation of cohesive ends on PCR products by UDG-mediated d. excision, U. of, and application for cloning into restriction digest-linearized vectors, *PCR Methods Appl.*, 1993, **2**, 328–32.
55 C. Aslanidis and P. J. de Jong, Ligation-independent cloning of PCR products (LIC-PCR), *Nucleic Acids Res.*, 1990, **18**, 6069–74.
56 C. Li and R. M. Evans, Ligation independent cloning irrespective of restriction site compatibility, *Nucleic Acids Res.*, 1997, **25**, 4165–6.
57 M. Li and S. Elledge, Harnessing homologous recombination *in vitro* to generate recombinant DNA *via* SLIC, *Nat. Methods*, 2007, **4**, 251.
58 D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nat. Methods*, 2009, **6**, 343–5.
59 D. G. Gibson, H. O. Smith, C. A. Hutchison, J. C. Venter and C. Merryman, Chemical synthesis of the mouse mitochondrial genome, *Nat. Methods*, 2010, **7**, 901–3.
60 P. A. Carr and G. M. Church, Genome engineering, *Nat. Biotechnol.*, 2009, **27**, 1151–62.
61 P. Montero L1opis, A. Jackson and O. Sliusarenko, *et al.* Spatial organization of the flow of genetic information in bacteria, *Nature*, 2010, **466**, 77.
62 D. Gibson, G. Benders and K. Axelrod, *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20404.
63 G. Benders, V. Noskov and E. Denisova, *et al.* Cloning whole bacterial genomes in yeast, *Nucleic Acids Res.*, 2010, **38**, 2558.
64 V. Larionov, N. Kouprina, J. Graves, X. N. Chen, J. R. Korenberg and M. A. Resnick, Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination, *Proc.Natl. Acad. Sci. U. S. A.*, 1996, **93**, 491–6.
65 P. Marschall, N. Malik and Z. Larin, Transfer of YACs up to 2.3 Mb intact into human cells with polyethylenimine, *Gene Ther.*, 1999, **6**, 1634–7.
66 V. N. Noskov, M. Koriabine and G. Solomon, *et al.* Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning, *Nucleic Acids Res.*, 2001, **29**, E32.
67 D. Gibson, Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides, *Nucleic Acids Res.*, 2009, **37**, 6984.
68 Z. Shao, H. Zhao and H. Zhao, DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways, *Nucleic Acids Res.*, 2009, **37**, e16.
69 M. Itaya, K. Fujita, A. Kuroki and K. Tsuge, Bottom-up genome assembly using the Bacillus subtilis genome vector, *Nat. Methods*, 2008, **5**, 41.
70 M. Itaya, K. Tsuge, M. Koizumi and K. Fujita, Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15971–6.
71 M. Itaya and S. Kaneko, Integration of stable extracellular DNA released from Escherichia coli into the Bacillus subtilis genome vector by culture mix method, *Nucleic Acids Res.*, 2010, **38**, 2551.
72 C. Lartigue, J. Glass and N. Alperovich, *et al.* Genome transplantation in bacteria: changing one species to another, *Science*, 2007, **317**, 632.
73 Y. Cai, B. Hartnett, C. Gustafsson and J. Peccoud, A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts, *Bioinformatics*, 2007, **23**, 2760–7.
74 D. Densmore, T. Hsiau, J. Kittleson, W. DeLoache, C. Batten and C. Anderson, Algorithms for automated DNA assembly, *Nucleic Acids Res.*, 2010, **38**, 2607.
75 ClothoCAD. D. Densmore, http://clothocad.sourceforge.net/wiki/index.php/ClothoCAD.
76 J5. N. Hillson, http://jbei-exwebapp.lbl.gov/j5/index.php/Main_Page.
77 H. H. Wang, F. J. Isaacs and P. A. Carr, *et al.* Programming cells by multiplex genome engineering and accelerated evolution, *Nature*, 2009, **460**, 894–8.